

STRUCTURE-PRESERVING SCHUR METHODS FOR COMPUTING SQUARE ROOTS OF REAL SKEW-HAMILTONIAN MATRICES *

ZHONGYUN LIU [†], YULIN ZHANG [‡], CARLA FERREIRA [‡], AND RUI RALHA [‡]

Abstract. Our contribution is two-folded. First, starting from the known fact that every real skew-Hamiltonian matrix has a real Hamiltonian square root, we give a complete characterization of the square roots of a real skew-Hamiltonian matrix W . Second, we propose a structure-exploiting method for computing square roots of W . Compared to the standard real Schur method, which ignores the structure, our method requires significantly less arithmetic.

Key words. Matrix square root, skew-Hamiltonian Schur decomposition, structure-preserving algorithm

AMS subject classifications. 65F15, 65F30, 15A18

1. Introduction. Given $A \in \mathbb{C}^{n \times n}$, a matrix X for which $X^2 = A$ is called a *square root* of A . The matrix square root is a useful theoretical and computational tool, one of the most commonly occurring matrix functions. See [10, 13, 15, 16, 17, 20].

The theory behind the existence of matrix square roots is nontrivial and the feature which complicates this theory is that in general not all the square roots of a matrix A are functions of A . See [6, 20].

It is well known that certain matrix structures can be inherited by the square root. For example, a symmetric positive (semi)definite matrix has a unique symmetric positive (semi)definite square root [19]. The square roots of a centrosymmetric matrix are also centrosymmetric [23]. A nonsingular M -matrix has exactly one M -matrix as a square root. For an H -matrix with positive diagonal elements there exists one and only one square root which is also an H -matrix with positive diagonal elements [21]. The principal square root of a centrosymmetric H -matrix with positive diagonal

* The first author is supported by the National Natural Science Foundations of China, N°.10771022 and N°.10571012; Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry. The research of the remaining authors is supported by FEDER Funds through “Programa Operacional Factores de Competitividade - COMPETE” and by Portuguese Funds through FCT - “Fundação para a Ciência e a Tecnologia”, within the Project PEst-CMAT/UI0013/2011.

[†]School of Math., Changsha University of Science & Technology, Hunan, 410076, China (liuzhongyun@263.net). Corresponding author.

[‡]Centro de Matemática, Universidade do Minho, 4710-057 Braga, Portugal (zhang@math.uminho.pt, caferrei@math.uminho.pt, r_ralha@math.uminho.pt).

elements is a unique centrosymmetric H -matrix with positive diagonal entries [22]. Any real skew-Hamiltonian matrix has a real Hamiltonian square root [8]. In this paper we characterize such square roots.

For general matrices, an attractive method which uses the Schur decomposition is described by Björck and Hammarling [4] but may require complex arithmetic. Higham [12] presented a modification of this method which enables real arithmetic to be used throughout when computing a real square root of a real matrix. This method has been extended to compute matrix p th roots [27] and general matrix functions [7].

It is a basic tenet in numerical analysis that structure should be exploited allowing, in general, the development of faster and/or more accurate algorithms [3, 25]. We propose a structure-exploiting method for computing square roots of a real skew-Hamiltonian matrix W which uses the real skew-Hamiltonian Schur decomposition and requires significantly less arithmetic.

We give some basic definitions and establish notation in Section 2. A description of the real Schur method and some results concerning the existence of real square roots are presented in Section 3. In Section 4 we characterize the square roots of a nonsingular W in a manner which makes clear the distinction between the square roots which are functions of W and those which are not. In Section 5 we present our algorithms for the computation of skew-Hamiltonian and Hamiltonian square roots. In Section 6 we give results of numerical experiments.

2. Definitions and preliminaries results.

2.1. Square roots of a nonsingular matrix. Given a scalar function f and a matrix $A \in \mathbb{C}^{n \times n}$ there are many different ways to define $f(A)$, a matrix of the same dimension of A , providing a useful generalization of a function of a scalar variable.

It is a standard result that any matrix $A \in \mathbb{C}^{n \times n}$ can be expressed in the Jordan canonical form

$$Z^{-1}AZ = J = \text{diag}(J_1, J_2, \dots, J_p), \quad (2.1)$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k} \quad (2.2)$$

where Z is nonsingular and $m_1 + m_2 + \dots + m_p = n$. The Jordan matrix J is unique up to the ordering of the blocks J_i .

Denote by $\lambda_1, \dots, \lambda_s$ the s distinct eigenvalues of A and let n_i be the order of the

largest Jordan block in which λ_i appears. The function f is said to be *defined on the spectrum of A* if the values

$$f^{(j)}(\lambda_i), \quad j = 0, \dots, n_i - 1, \quad i = 1, \dots, s,$$

exist. These are called the *values of the function f on the spectrum of A* .

The following definition of matrix function defines $f(A)$ to be a polynomial in the matrix A completely determined by the values of f on the spectrum of A . See [12, p. 407 ff.].

DEFINITION 2.1 (matrix function via Hermite interpolation). Let f be defined on the spectrum of $A \in \mathbb{C}^{n \times n}$. Then

$$f(A) := p(A)$$

where p is the polynomial of degree less than $\sum_1^s n_i$ which satisfies the interpolation conditions

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), \quad j = 0, \dots, n_i - 1, \quad i = 1, \dots, s.$$

There is a unique such p and it is known as the Hermite interpolating polynomial.

Of particular interest here is the function $f(z) = z^{1/2}$ which is certainly defined on the spectrum of A if A is nonsingular. However, the square root function of A , $f(A)$, is not uniquely defined until one specifies which branch of the square root is to be taken in the neighborhood of each eigenvalue λ_i . Indeed, Definition 2.1 yields a total of 2^s matrices $f(A)$ when all combinations of branches for the square roots $f(\lambda_i)$, $i = 1, \dots, s$, are taken. It is natural to ask whether these matrices are in fact square roots of A , that is, do we have $f(A)f(A) = A$? Indeed, these matrices, which are polynomials in A by definition, are square roots of A . See [15, 20]. However, these square roots are not necessarily all the square roots of A .

To classify all the square roots of a nonsingular matrix $A \in \mathbb{C}^{n \times n}$ we need the following result concerning the square roots of a Jordan block.

LEMMA 2.2. For $\lambda_k \neq 0$ the Jordan block $J_k(\lambda_k)$ in (2.2) has precisely two upper triangular square roots

$$L_k^{(j)} = L_k^{(j)}(\lambda_k) = \begin{bmatrix} f(\lambda_k) & f'(\lambda_k) & \cdot & \dots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & f'(\lambda_k) & \dots & \frac{f^{(m_k-2)}(\lambda_k)}{(m_k-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & f(\lambda_k) & f'(\lambda_k) \\ & & & & f(\lambda_k) \end{bmatrix}, \quad j = 1, 2, \quad (2.3)$$

where $f(\lambda) = \lambda^{1/2}$ and the superscript j denotes the branch of the square root in the neighborhood of λ_k . Both square roots are functions of J_k .

We will restrict our attention to matrices with real entries and to investigate the real square roots of a real matrix we need to understand the structure of a general complex square root. The following results allow us to obtain a useful characterisation of the square roots of a nonsingular matrix $A \in \mathbb{C}^{n \times n}$ which are functions of A . See [10, 12].

THEOREM 2.3. *Let $A \in \mathbb{C}^{n \times n}$ be nonsingular and have the Jordan canonical form (2.2). Then all square roots X of A are given by*

$$X = ZU \operatorname{diag} \left(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)} \right) U^{-1} Z^{-1},$$

where $j_k = 1$ or $j_k = 2$ and U is an arbitrary nonsingular matrix which commutes with J .

The following result extends Theorem 2.3.

THEOREM 2.4. *Let the nonsingular matrix $A \in \mathbb{C}^{n \times n}$ have the Jordan canonical form (2.2) and let $s \leq p$ be the number of distinct eigenvalues of A . Then A has precisely 2^s square roots which are functions of A , given by*

$$X_j = Z \operatorname{diag} \left(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)} \right) Z^{-1}, \quad j = 1, \dots, 2^s, \quad (2.4)$$

corresponding to all possible choices of j_1, \dots, j_p , $j_k = 1$ or $j_k = 2$, subject to the constraint that $j_i = j_k$ whenever $\lambda_i = \lambda_k$.

If $s < p$, A has square roots which are not functions of A ; they form parametrized families

$$X_j(U) = ZU \operatorname{diag} \left(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_p^{(j_p)} \right) U^{-1} Z^{-1}, \quad j = 2^s + 1, \dots, 2^p, \quad (2.5)$$

where $j_k = 1$ or $j_k = 2$, U is an arbitrary nonsingular matrix which commutes with J , and for each j there exist i and k , depending on j , such that $\lambda_i = \lambda_k$ while $j_i \neq j_k$.

Proofs of these theorems and a description of the structure of the matrix U can be found in [10]. Note that formula in (2.4) follows from the fact that all square roots of A which are functions of A have the form

$$f(A) = f(ZJZ^{-1}) = Zf(J)Z^{-1} = Z \operatorname{diag} (f(J_k)) Z^{-1},$$

and from Lemma 2.2. The constrain on the branches $\{j_i\}$ follow from Definition 2.1. The remaining square roots of A (if any), which cannot be functions of A , are given by (2.5).

Theorem 2.4 shows that the square roots of A which are functions of A are “isolated” square roots, characterized by the fact that the sum of any two of their eigenvalues is nonzero. On the other hand, the square roots which are not functions of A form a finite number of parametrized families of matrices: each family contains infinitely many square roots which share the same spectrum.

Some interesting corollaries follow directly from Theorem 2.4.

COROLLARY 2.5. *If $\lambda_k \neq 0$, the two square roots of $J_k(\lambda_k)$ given in Lemma 2.2 are the only square roots of $J_k(\lambda_k)$.*

COROLLARY 2.6. *If $A \in \mathbb{C}^{n \times n}$ is nonsingular and in its Jordan canonical form (2.2) each eigenvalue appears in only one Jordan block, then A has precisely 2^p square roots, each of which is a function of A .*

The final corollary is well known.

COROLLARY 2.7. *Every Hermitian positive definite matrix has a unique Hermitian positive definite square root.*

2.2. Hamiltonian and skew-Hamiltonian matrices. Hamiltonian and skew-Hamiltonian matrices have properties that follow directly from the definition.

DEFINITION 2.8. Let $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$, where I is the identity matrix of order n .

- (1) A matrix $H \in \mathbb{R}^{2n \times 2n}$ is said to be *Hamiltonian* if $HJ = (HJ)^T$. Equivalently, H can be partitioned as

$$H = \begin{bmatrix} A & G \\ F & -A^T \end{bmatrix}, \quad G = G^T, \quad F = F^T, \quad A, G, F \in \mathbb{R}^{n \times n}. \quad (2.6)$$

- (2) A matrix $W \in \mathbb{R}^{2n \times 2n}$ is said to be *skew-Hamiltonian* if $WJ = -(WJ)^T$. Likewise, W can be partitioned as

$$W = \begin{bmatrix} A & G \\ F & A^T \end{bmatrix}, \quad G = -G^T, \quad F = -F^T, \quad A, G, F \in \mathbb{R}^{n \times n}. \quad (2.7)$$

These matrix structures induce particular spectral properties for H and W . Notably, the eigenvalues of H are symmetric with respect to the imaginary axis and the eigenvalues of W have even algebraic and geometric multiplicities.

DEFINITION 2.9.

- (1) A matrix $S \in \mathbb{R}^{2n \times 2n}$ is said to be *symplectic* if $SJS^T = J$.
 (2) A matrix $U \in \mathbb{R}^{2n \times 2n}$ is said to be *orthogonal-symplectic* if U is orthogonal and symplectic. Any matrix belonging to this group can be par-

tioned as

$$U = \begin{bmatrix} U_1 & U_2 \\ -U_2 & U_1 \end{bmatrix}$$

where $U_i \in \mathbb{R}^{n \times n}$, $i = 1, 2$.

Hamiltonian and skew-Hamiltonian structures are preserved if symplectic similarity transformations are used; if H is Hamiltonian (skew-Hamiltonian) and S is symplectic, then $S^{-1}HS$ is also Hamiltonian (skew-Hamiltonian). In the interest of numerical stability the similarities should be orthogonal as well.

The first simplifying reduction of a skew-Hamiltonian matrix was introduced by Van Loan in [24]. But first we recall the real Schur decomposition [11].

THEOREM 2.10 (real Schur form). *If $A \in \mathbb{R}^{n \times n}$, then there exists a real orthogonal matrix Q such that*

$$Q^T A Q = R = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1m} \\ & R_{22} & \dots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad (2.8)$$

where each block R_{ii} is either 1×1 or 2×2 with complex conjugate eigenvalues λ_i and $\bar{\lambda}_i$, $\lambda_i \neq \bar{\lambda}_i$ (R is in quasi-upper triangular form).

In [24] it was shown that any skew-Hamiltonian W can be brought to block-upper-triangular form by an orthogonal-symplectic similarity. Actually, we can explicitly compute an orthogonal-symplectic matrix U such that

$$U^T W U = \begin{bmatrix} W_1 & W_2 \\ O & W_1^T \end{bmatrix},$$

where $W_2^T = -W_2$ and W_1 is upper Hessenberg (a matrix is *upper Hessenberg* if all entries below its first subdiagonal are zero). This is called the *symplectic Paige/Van Loan (PVL) form*.

Subsequently, if the standard QR algorithm is applied to W_1 producing an orthogonal matrix Q and a matrix in real Schur form N_1 so that

$$W_1 = Q N_1 Q^T,$$

we attain the *real skew-Hamiltonian Schur decomposition* of W ,

$$\mathcal{U}^T W \mathcal{U} = \begin{bmatrix} N_1 & N_2 \\ O & N_1^T \end{bmatrix}, \quad (2.9)$$

where $\mathcal{U} = U \begin{bmatrix} Q & O \\ O & Q \end{bmatrix}$ and $N_2 = Q^T W_2 Q$.

LEMMA 2.11 (real skew-Hamiltonian Schur form). *Let $W \in \mathbb{R}^{2n \times 2n}$ be skew-Hamiltonian. Then there exists an orthogonal matrix*

$$\mathcal{U} = \begin{bmatrix} \mathcal{U}_1 & \mathcal{U}_2 \\ -\mathcal{U}_2 & \mathcal{U}_1 \end{bmatrix}, \quad \mathcal{U}_1, \mathcal{U}_2 \in \mathbb{R}^{n \times n},$$

such that

$$\mathcal{U}^T W \mathcal{U} = \begin{bmatrix} N_1 & N_2 \\ 0 & N_1^T \end{bmatrix}, \quad N_2^T = -N_2, \quad (2.10)$$

and N_1 is in real Schur form.

In [26, Theorem 5.1] we can find a result concerning the *real Hamiltonian Schur decomposition*.

THEOREM 2.12 (real Hamiltonian Schur form). *Let $H \in \mathbb{R}^{2n \times 2n}$ be Hamiltonian. If H has no nonzero purely imaginary eigenvalues, then there exists an orthogonal matrix*

$$\mathcal{U} = \begin{bmatrix} \mathcal{U}_1 & \mathcal{U}_2 \\ -\mathcal{U}_2 & \mathcal{U}_1 \end{bmatrix}, \quad \mathcal{U}_1, \mathcal{U}_2 \in \mathbb{R}^{n \times n},$$

such that

$$\mathcal{U}^T H \mathcal{U} = \begin{bmatrix} H_1 & H_2 \\ 0 & -H_1^T \end{bmatrix}, \quad H_2^T = H_2, \quad (2.11)$$

and H_1 is in real Schur form.

In this article we are interested in the computation of a real square root of a real skew-Hamiltonian matrix W and to discuss square roots of a skew-Hamiltonian matrix we need to consider a variant of the Jordan canonical form (2.2) when the matrix A is real. In this case, all the nonreal eigenvalues must occur in conjugate pairs and all the Jordan blocks of all sizes (not just 1×1 blocks) corresponding to nonreal eigenvalues occur in conjugate pairs of equal size.

For example, if λ is a nonreal eigenvalue of the real matrix A , and if $J_2(\lambda)$ appears in the Jordan canonical form of A with a certain multiplicity, $J_2(\bar{\lambda})$ must also appear with the same multiplicity. See [18, p.150 ff.]. The block matrix

$$\begin{bmatrix} J_2(\lambda) & O \\ O & J_2(\bar{\lambda}) \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ \hline 0 & 0 & \bar{\lambda} & 1 \\ 0 & 0 & 0 & \bar{\lambda} \end{bmatrix} \quad (2.12)$$

is permutation-similar (interchange rows and columns 2 and 3) to the block matrix

$$\begin{bmatrix} \lambda & 0 & 1 & 0 \\ 0 & \bar{\lambda} & 0 & 1 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \bar{\lambda} \end{bmatrix} = \begin{bmatrix} D(\lambda) & I \\ 0 & D(\lambda) \end{bmatrix}, \quad D(\lambda) := \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}. \quad (2.13)$$

Each block $D(\lambda)$ is similar to the matrix

$$SD(\lambda)S^{-1} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} := C(a, b), \quad S = \begin{bmatrix} -i & -i \\ 1 & -1 \end{bmatrix}, \quad (2.14)$$

where $\lambda, \bar{\lambda} = a \pm ib$, $a, b \in \mathbb{R}$, $b \neq 0$. Thus, every block pair of conjugate 2×2 Jordan blocks (2.12) with nonreal eigenvalue λ is similar to a real 4×4 block of the form

$$\begin{bmatrix} a & b & 1 & 0 \\ -b & a & 0 & 1 \\ 0 & 0 & a & b \\ 0 & 0 & -b & a \end{bmatrix} = \begin{bmatrix} C(a, b) & I \\ O & C(a, b) \end{bmatrix}.$$

In general, every block pair of conjugate $k \times k$ Jordan blocks with nonreal λ ,

$$\begin{bmatrix} J_k(\lambda) & O \\ O & J_k(\bar{\lambda}) \end{bmatrix}, \quad (2.15)$$

is similar to a real $2k \times 2k$ block matrix of the form

$$C_k(a, b) = \begin{bmatrix} C(a, b) & I & & & \\ & C(a, b) & I & & \\ & & \ddots & \ddots & \\ & & & C(a, b) & I \\ & & & & C(a, b) \end{bmatrix}. \quad (2.16)$$

We call $C_k(a, b)$ a *real Jordan block*. These observations lead us to the *real Jordan canonical form*.

THEOREM 2.13. *Each matrix $A \in \mathbb{R}^{n \times n}$ is similar (via a real similarity transformation) to a block diagonal real matrix of the form*

$$J_R = \begin{bmatrix} C_{n_1}(a_1, b_1) & & & & \\ & \ddots & & & \\ & & C_{n_p}(a_p, b_p) & & \\ & & & J_{n_{p+1}}(\lambda_{p+1}) & \\ & & & & \ddots \\ & & & & & J_{n_{p+q}}(\lambda_{p+q}) \end{bmatrix}, \quad (2.17)$$

where $\lambda_k = a_k + ib_k$, $a_k, b_k \in \mathbb{R}$, $k = 1, \dots, p$, is a nonreal eigenvalue of A and λ_k , $k = p + 1, \dots, p + q$, is a real eigenvalue of A . Each real Jordan block $C_{n_k}(a_k, b_k)$ is of the form (2.16) and corresponds to a pair of conjugate Jordan blocks $J_{n_k}(\lambda_k)$ and $J_{n_k}(\bar{\lambda}_k)$ for a nonreal λ_k in the Jordan canonical form of A (2.1). The real Jordan blocks $J_{n_k}(\lambda_k)$ are exactly the Jordan blocks in (2.1) with real λ_k . Notice that $2(n_1 + \dots + n_p) + (n_{p+1} + \dots + n_{p+q}) = n$. We call J_R a real Jordan matrix of order n , a direct sum of real Jordan blocks.

In [8] it is shown that every real skew-Hamiltonian matrix can also be reduced to a real skew-Hamiltonian Jordan form via a symplectic similarity. See also [9].

LEMMA 2.14. [8, Theorem 1] For every real skew-Hamiltonian matrix $W \in \mathbb{R}^{2n \times 2n}$ there exists a symplectic matrix $\Psi \in \mathbb{R}^{2n \times 2n}$ such that

$$\Psi^{-1}W\Psi = \begin{bmatrix} J_R & \\ & J_R^T \end{bmatrix}, \quad (2.18)$$

where $J_R \in \mathbb{R}^{n \times n}$ is in real Jordan form (2.17) and is unique up to a permutation of real Jordan blocks.

In principle, a real square root of W can be obtained by the general method devised by Higham [12]. Such method, however, does not exploit the structure of W . The method we propose exploits the skew-Hamiltonian structure of W and it also uses the real Schur method. To make our presentation simpler we decided to present the details of the real Schur method. See [12, p. 412 ff.].

3. An algorithm for computing real square roots.

3.1. The Schur method. Björck and Hammarling [4] presented an excellent method for computing a square root of a matrix A . Their method first computes a Schur decomposition

$$Q^*AQ = T$$

where Q is unitary and T is upper triangular [11], and then determines an upper triangular square root Y of T with the aid of a fast recursion. A square root of A is given by

$$X = QYQ^*.$$

A disadvantage of this Schur method is that if A is real and has nonreal eigenvalues, the method needs complex arithmetic even if the square root which is computed should be real. When computing a real square root it is obviously desirable to work with real arithmetic; depending on the relative costs of real and complex arithmetic on a

given computer system, substantial computational savings may occur, and moreover, a computed real square root is guaranteed.

Higham described a generalization of the Schur method which enables real arithmetic to be used throughout when computing a real square root of a real matrix. In Section 3.3 we present this method. First we give, for a nonsingular matrix $A \in \mathbb{R}^{n \times n}$, conditions for the existence of a real square root, and for the existence of a real square root which is a polynomial in A .

3.2. Existence of real square roots. The following result concerns the existence of general real square roots - those which are not necessarily functions of A .

THEOREM 3.1. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular. A has a real square root if and only if each elementary divisor of A corresponding to a real negative eigenvalue occurs an even number of times.*

Theorem 3.1 is mainly of theoretical interest, since the proof is nonconstructive and the condition for the existence of a real square root is not easily checked computationally. We now focus attention on the real square roots of $A \in \mathbb{R}^{n \times n}$. The key to analysing the existence of square roots of this type is the real Schur decomposition.

Suppose that $A \in \mathbb{R}^{n \times n}$ and that f is defined on the spectrum of A and consider the real Schur form of A in (2.8). Since A and R in (2.8) are similar, we have

$$f(A) = Qf(R)Q^T,$$

so that $f(A)$ is real if and only if

$$Z = f(R)$$

is real. It is not difficult to show that Z inherits R 's quasi-upper triangular structure and that

$$Z_{ii} = f(R_{ii}), \quad i = 1, \dots, m.$$

If A is nonsingular and f is the square root function, then we have

$$Z^2 = R \quad \text{and} \quad X^2 = A \quad \text{with} \quad X = QZQ^T.$$

The whole of Z is uniquely determined by its diagonal blocks. To see this equate (i, j) blocks in the equation $Z^2 = R$ to obtain

$$\sum_{k=i}^j Z_{ik}Z_{kj} = R_{ij}, \quad j \geq i.$$

These equations can be recast in the form

$$Z_{ii}^2 = R_{ii}, \quad i = 1, \dots, m \quad (3.1)$$

$$Z_{ii}Z_{ij} + Z_{ij}Z_{jj} = R_{ij} - \sum_{k=i+1}^{j-1} Z_{ik}Z_{kj}, \quad j = i+1, \dots, m. \quad (3.2)$$

Thus, if the diagonal blocks Z_{ii} are known, (3.1) provides an algorithm for computing the remaining blocks Z_{ij} of Z along one superdiagonal at a time in the order specified by $j - i = 1, 2, \dots, m - 1$. The condition for the Sylvester equation (3.2) to have a unique solution Z_{ij} is that Z_{ii} and $-Z_{jj}$ have no eigenvalue in common [11, 20]. This is guaranteed because the eigenvalues of Z are $\mu_k = f(\lambda_k)$ and for the square root function $f(\lambda_i) = -f(\lambda_j)$ implies that $\lambda_i = \lambda_j = 0$, contradicting the nonsingularity of A .

From this algorithm for constructing Z from its diagonal blocks we conclude that Z is real and hence $f(A)$ is real, if and only if each of the blocks $Z_{ii} = f(R_{ii})$ is real.

We now examine the square roots $f(A)$ of a 2×2 matrix A with complex conjugate eigenvalues. Since A has 2 distinct eigenvalues, it follows from Corollary 2.6 that A has four square roots which are all functions of A . Next lemma says about the form of these square roots.

LEMMA 3.2. *Let $A \in \mathbb{R}^{2 \times 2}$ have complex eigenvalues $\lambda, \bar{\lambda} = \theta \pm i\mu$, where $\mu \neq 0$. Then A has four square roots, each of which is a function of A . Two of the square roots are real, with complex conjugate eigenvalues, and two are pure imaginary, having eigenvalues which are not complex conjugate.*

Using this lemma it can be proved

THEOREM 3.3. *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular. If A has a real negative eigenvalue, then A has no real square roots which are functions of A .*

If A has no real negative eigenvalues, then there are precisely 2^{r+c} real square roots of A which are functions of A , where r is the number of distinct real eigenvalues of A and c is the number of distinct complex conjugate eigenvalue pairs.

For proofs of Lemma 3.2 and Theorem 3.3 see [12, p. 414 ff.].

It is clear from Theorem 3.1 that A may have real negative eigenvalues and yet still have a real square root; however, as Theorem 3.3 shows, the square root will not be a function of A .

3.3. The real Schur method. The ideas of the last section lead to a natural extension of Björck and Hammarling's Schur method for computing in real arithmetic a real square root of a nonsingular $A \in \mathbb{R}^{n \times n}$. This real Schur method begins by

computing a real Schur decomposition (2.8), then computes a square root Z of R from equations (3.1) and (3.2), and finally obtains a square root of A via the transformation $X = QZQ^T$.

The solution of Equation (3.1) can be computed efficiently in a way suggested by the proof of Lemma 3.2 [12, p. 417]. The first step is to compute θ and μ , where $\lambda = \theta + i\mu$ is an eigenvalue of the matrix

$$R_{ii} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}.$$

Next, α and β such that $(\alpha + i\beta)^2 = \theta + i\mu$ are required. Finally, the real square roots of R_{ii} are given by

$$Z_{ii} = \pm \left(\alpha I + \frac{1}{2\alpha} (R_{ii} - \theta I) \right). \quad (3.3)$$

If Z_{ii} is of order p and Z_{jj} is of order q , Equation (3.2) can be written as

$$(I_q \otimes Z_{ii} + Z_{jj}^T \otimes I_p) \operatorname{col}(Z_{ij}) = \operatorname{col} \left(R_{ij} - \sum_{k=i+1}^{j-1} Z_{ik} Z_{kj} \right), \quad j > i \quad (3.4)$$

where \otimes is the Kronecker product and $\operatorname{col}(M)$ denotes the column vector formed by taking columns of M and stacking them atop one another from left to right. The linear system (3.4) is of order $pq = 1, 2$ or 4 and may be solved by standard methods.

Note that to conform with the definition of $f(A)$ we have to choose the signs in (3.3) so that Z_{ii} and Z_{jj} have the same eigenvalues whenever R_{ii} and R_{jj} do; this choice ensures simultaneously the nonsingularity of the linear systems (3.4).

Any of the real square roots $f(A)$ of A can be computed in the above fashion by the real Schur method.

Algorithm 1 [Real Schur method]

1. compute a real Schur decomposition of A ,

$$A = Q^T R Q;$$

2. compute a square root Z of R solving the equation $Z^2 = R$ via

$$Z_{ii}^2 = R_{ii}, \quad 1 \leq i \leq m,$$

$$(I_q \otimes Z_{ii} + Z_{jj}^T \otimes I_p) \operatorname{col}(Z_{ij}) = \operatorname{col} \left(R_{ij} - \sum_{k=i+1}^{j-1} Z_{ik} Z_{kj} \right), \quad j > i$$

[block fast recursion]

3. obtain a square root of A , $X = QZQ^T$.

The cost of the real Schur method, measured in floating point operations (flops) may be broken down as follows. The real Schur factorization costs about $15n^3$ flops [11]. The computation of Z requires $n^3/6$ flops and the formation of $X = QZQ^T$ requires $3n^3/2$ flops [12, p. 418]. Only a fraction of the overall time is spent in computing the square root Z .

4. Square roots of a skew-Hamiltonian matrix. In this section we present a detailed classification of the square roots of a skew-Hamiltonian matrix $W \in \mathbb{R}^{2n \times 2n}$ based on its real skew-Hamiltonian Jordan form (2.18),

$$\Psi^{-1}W\Psi = \begin{bmatrix} J_R & \\ & J_R^T \end{bmatrix}, \quad (4.1)$$

where $J_R \in \mathbb{R}^{n \times n}$ is in real Jordan form (2.17) and Ψ is a symplectic matrix. First, we will discuss the square roots of J_R .

According to Lemma 2.2, for $\lambda_k \neq 0$ a canonical Jordan block $J_k(\lambda_k)$ has precisely two upper triangular square roots given by (2.3). As a corollary it follows that

COROLLARY 4.1. *For a real eigenvalue $\lambda_k \neq 0$, the Jordan block $J_k(\lambda_k)$ in (2.2) has precisely two upper triangular square roots which are real, if $\lambda > 0$, and pure imaginary, if $\lambda < 0$. Both square roots are functions of J_k .*

To fully characterize the square roots of a real Jordan block $C_k(a, b)$, we first examine the square roots of a 2×2 block $C(a, b)$ corresponding to the nonreal eigenvalues $\lambda, \bar{\lambda} = a \pm ib$, $a, b \in \mathbb{R}$. According to Lemma 3.2, the real Jordan block $C(a, b)$ in (2.14) has four square roots, each of which is a function of $C(a, b)$. Two of the square roots are real and two are pure imaginary.

LEMMA 4.2. *The real Jordan block $C_k(a, b)$ in (2.16) has precisely four block upper triangular square roots*

$$F_k^{(j)} = \begin{bmatrix} F & F_1 & \cdot & \dots & F_{k-1} \\ & F & F_1 & \dots & F_{k-2} \\ & & \ddots & \ddots & \vdots \\ & & & F & F_1 \\ & & & & F \end{bmatrix}, \quad j = 1, \dots, 4, \quad (4.2)$$

where F is a square root of $C(a, b)$ and F_i , $i = 1, \dots, k-1$, are the unique solutions of certain Sylvester equations. The superscript j denotes one of the four square roots of $C(a, b)$. These four square roots $F_k^{(j)}$ are functions of $C_k(a, b)$, two of them are real and two are pure imaginary.

Proof. Since $C_k(a, b)$ has 2 distinct eigenvalues and the Jordan form (2.15) has

$p = 2$ blocks, from Corollary 2.6 we know that $C_k(a, b)$ has four square roots which are all functions of A .

Let X be a square root of $C_k(a, b)$ ($k > 1$). It is not difficult to see that X inherits $C_k(a, b)$ block upper triangular structure,

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1,k} \\ & X_{22} & \cdots & X_{2,k} \\ & & \ddots & \vdots \\ & & & X_{kk} \end{bmatrix} \quad (4.3)$$

where $X_{i,j}$ are all 2×2 matrices. Equating (i, j) blocks in the equation

$$X^2 = C_k(a, b)$$

we obtain

$$X_{ii}^2 = C(a, b), \quad i = 1, \dots, k, \quad (4.4)$$

$$X_{ii}X_{i,i+1} + X_{i,i+1}X_{i+1,i+1} = I_2, \quad i = 1, \dots, k-1 \quad (4.5)$$

$$X_{ii}X_{ij} + X_{ij}X_{jj} = - \sum_{l=i+1}^{j-1} X_{il}X_{lj}, \quad j = i+2, \dots, k. \quad (4.6)$$

The whole of X is uniquely determined by its diagonal blocks. If F is one square root of $C(a, b)$, from (4.4) and to conform with the definition of $f(A)$ (the eigenvalues of X_{ii} must be the same), we have

$$X_{ii} = F, \quad i = 1, \dots, k. \quad (4.7)$$

Equations (4.5) and (4.6) are Sylvester equations and the condition for them to have a unique solution X_{ij} is that X_{ii} and $-X_{jj}$ have no eigenvalues in common and this is guaranteed.

From (4.5) we obtain the blocks X_{ij} along the first superdiagonal and (4.7) forces them to be all equal, say F_1 ,

$$X_{12} = X_{23} = \dots = X_{k-1,k} = F_1.$$

This implies that the other superdiagonals obtained from (4.6) are also constant, say F_{j-1} , $j = 3, \dots, k$,

$$X_{1j} = X_{2,j+1} = \dots = X_{k-j+1,k} = F_{j-1}, \quad j = 3, \dots, k.$$

Thus, since there are only exactly four distinct square roots of $C(a, b)$ which are functions of $C(a, b)$, $F = F^{(l)}$, $l = 1, \dots, 4$, it follows that $C_k(a, b)$ will also have

precisely four square roots which are functions of $C_k(a, b)$. If F is real then F_{j-1} , $j = 2, \dots, k$, will also be real. If F is pure imaginary it can also be seen that F_{j-1} , $j = 2, \dots, k$, will be pure imaginary too. \square

Next theorem combines Corollary 4.1 and Lemma 4.2 to characterize the square roots of a real Jordan matrix J_R .

THEOREM 4.3. *Assume that a nonsingular real Jordan matrix J_R in (2.17) has p real Jordan blocks corresponding to c distinct complex conjugate eigenvalue pairs and q canonical Jordan blocks corresponding to r distinct real eigenvalues.*

Then J_R has precisely 2^{2c+r} square roots which are functions of J_R , given by

$$X_j = \text{diag} \left(F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}, L_{n_{p+1}}^{(i_1)}, \dots, L_{n_{p+q}}^{(i_q)} \right), \quad j = 1, \dots, 2^{2c+r}, \quad (4.8)$$

corresponding to all possible choices of j_1, \dots, j_p , $j_k = 1, 2, 3$ or 4 , and i_1, \dots, i_q , $i_k = 1$ or 2 , subject to the constraint that $j_l = j_k$ and $i_l = i_k$ whenever $\lambda_l = \lambda_k$.

If $c + r < p + q$, then J_R has square roots which are not functions of J_R and they form $2^{2p+q} - 2^{2c+r}$ parameterized families given by

$$X_j(\Omega) = \Omega \text{diag} \left(F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}, L_{n_{p+1}}^{(i_1)}, \dots, L_{n_{p+q}}^{(i_q)} \right) \Omega^{-1}, \quad (4.9)$$

$$j = 2^{2c+r} + 1, \dots, 2^{2p+q},$$

where $j_k = 1, 2, 3$ or 4 and $i_k = 1$ or 2 , Ω is an arbitrary nonsingular matrix which commutes with J_R and for each j there exist l and k depending on j , such that $\lambda_l = \lambda_k$ while $j_l \neq j_k$ or $i_l \neq i_k$.

Proof. The number of distinct eigenvalues is $s = 2c + r$ and, according to Theorem 2.4, J_R has precisely $2^s = 2^{2c+r}$ square roots which are functions of J_R . All square roots of J_R which are functions of J_R satisfy

$$f(J_R) = \begin{bmatrix} f(C_{n_1}) & & & & \\ & \ddots & & & \\ & & f(C_{n_p}) & & \\ & & & f(J_{n_{p+1}}) & \\ & & & & \ddots \\ & & & & & f(J_{n_{p+q}}) \end{bmatrix}$$

and, according to Lemma 2.2 and Lemma 4.2 these are given by (4.8). The constraint on the branches $\{j_k\}$ and $\{i_k\}$ comes from Definition 2.1 of matrix function. The remaining square roots of J_R , if they exist, cannot be functions of J_R . Equation (4.9) derives from the second part of Theorem 2.4. \square

From Theorem 4.3, Lemma 4.2 and Corollary 4.1 the next result concerning the square roots of A which are functions of A follows immediately.

COROLLARY 4.4. *Under the assumptions of Theorem 4.3,*

- (1) *if J_R has a real negative eigenvalue, then J_R has no real square roots which are functions of J_R ;*
- (2) *if J_R has no real negative eigenvalues, then J_R has precisely 2^{c+r} real square roots which are functions of J_R , given by (4.8) with the choices of j_1, \dots, j_p corresponding to real square roots $F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}$;*
- (3) *if J_R has no real positive eigenvalues, then J_R has precisely 2^{c+r} pure imaginary square roots which are functions of J_R , given by (4.8) with the choices of j_1, \dots, j_p corresponding to pure imaginary square roots $F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}$.*

Now we want to use all these results to characterize the square roots of a real skew-Hamiltonian matrix W .

THEOREM 4.5. *Let $W \in \mathbb{R}^{2n \times 2n}$ be a nonsingular skew-Hamiltonian matrix with the real skew-Hamiltonian Jordan form in (4.1). Assume that J_R has p real Jordan blocks corresponding to c distinct complex conjugate eigenvalue pairs and q canonical Jordan blocks corresponding to r distinct real eigenvalues.*

Then W has precisely 2^{2c+r} square roots which are functions of W , given by

$$Y_j = \Psi \operatorname{diag}(X_j, X_j^T) \Psi^{-1}, \quad j = 1, \dots, 2^{2c+r}, \quad (4.10)$$

where X_j is a square root of J_R given in (4.8).

W has always square roots which are not functions of W and they form $4^{2p+q} - 2^{2c+r}$ parameterized families given by

$$Y_j(\Theta) = \Psi \Theta \operatorname{diag}(\tilde{X}_j, \hat{X}_j^T) \Theta^{-1} \Psi^{-1}, \quad j = 2^{2c+r} + 1, \dots, 4^{2p+q}, \quad (4.11)$$

where

$$\begin{aligned} \tilde{X}_j &= \operatorname{diag}\left(F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}, L_{n_{p+1}}^{(i_1)}, \dots, L_{n_{p+q}}^{(i_q)}\right), \\ \hat{X}_j &= \operatorname{diag}\left(F_{n_1}^{(j_{p+1})}, \dots, F_{n_p}^{(j_{2p})}, L_{n_{p+q}}^{(i_{q+1})}, \dots, L_{n_{p+q}}^{(i_{2q})}\right), \end{aligned}$$

$j_k = 1, 2, 3$ or 4 and $i_k = 1$ or 2 , Θ is an arbitrary nonsingular matrix which commutes with $\operatorname{diag}(J_R, J_R^T)$ and for each j there exist l and k depending on j , such that $\lambda_l = \lambda_k$ while $j_l \neq j_k$ or $i_l \neq i_k$.

Notice that W has $s = 2c + r$ distinct eigenvalues corresponding to $2(p + q)$ real Jordan blocks and $2(2p + q)$ canonical Jordan blocks. We always have $s \leq 2p + q$ and so $s < 2(2p + q)$. Thus, there are always square roots which are not functions of W .

Proof. This result is a direct consequence of Theorem 2.4 and Theorem 4.3. Notice that if X_j in (4.8) is a square root of J_R then $\text{diag}(X_j, X_j^T)$ is a square root of $\text{diag}(J_R, J_R^T)$. Thus,

$$\begin{aligned} W &= \Psi \begin{bmatrix} J_R & \\ & J_R^T \end{bmatrix} \Psi^{-1} \\ &= \Psi \begin{bmatrix} X_j & \\ & X_j^T \end{bmatrix} \begin{bmatrix} X_j & \\ & X_j^T \end{bmatrix} \Psi^{-1} \\ &= \Psi \begin{bmatrix} X_j & \\ & X_j^T \end{bmatrix} \Psi^{-1} \Psi \begin{bmatrix} X_j & \\ & X_j^T \end{bmatrix} \Psi^{-1}. \end{aligned}$$

Thus,

$$Y_j = \Psi \begin{bmatrix} X_j & \\ & X_j^T \end{bmatrix} \Psi^{-1}$$

is a square root of W . Since X_j is a function of J_R , Y_j is a function of $\text{diag}(J_R, J_R^T)$. This proves the first part of the theorem.

The second part follows from the second part of Theorem 2.4 and the fact that $\text{diag}(\tilde{X}_j, \tilde{X}_j^T)$ is a square root of $\text{diag}(J_R, J_R^T)$. \square

It is easy to verify that if W is a skew-Hamiltonian matrix, then W^2 is also skew-Hamiltonian. This implies that any function of W , which is a polynomial by definition, is a skew-Hamiltonian matrix. Thus, all the square roots of W which are functions of W are skew-Hamiltonian matrices. The following result refers to the existence of real square roots of a skew-Hamiltonian matrix.

COROLLARY 4.6. *Under the assumptions of Theorem 4.5, the following statements hold:*

- (1) *if W has a real negative eigenvalue, then W has no real skew-Hamiltonian square roots;*
- (2) *if W has no real negative eigenvalues, then W has precisely 2^{c+r} real skew-Hamiltonian square roots which are functions of W , given by (4.10) with the choices of j_1, \dots, j_p corresponding to real square roots $F_{n_1}^{(j_1)}, \dots, F_{n_p}^{(j_p)}$.*

It is clear from Theorem 3.1 that W may have real negative eigenvalues and yet still have a real square root; however, the square root will not be a function of W .

In [8, Theorem 2] it is shown that

LEMMA 4.7. *Every real skew-Hamiltonian matrix W has a real Hamiltonian square root.*

The proof is constructive and the key step is based in Lemma 2.14 - we can bring W into a real skew-Hamiltonian Jordan form (4.1) via a symplectic similarity. Further, it is shown that every skew-Hamiltonian matrix W has infinitely many real Hamiltonian square roots.

The following theorem gives the structure of those real Hamiltonian square roots.

THEOREM 4.8. *Let $W \in \mathbb{R}^{2n \times 2n}$ be a nonsingular skew-Hamiltonian matrix and assume the conditions in Theorem 4.5.*

- (1) *If W has no real negative eigenvalues, then W has real Hamiltonian square roots which are not functions of W and they form 2^{p+q} parameterized families given by*

$$Y_j(\Theta) = \Psi \Theta \operatorname{diag}(X_j, -X_j^T) \Theta^{-1} \Psi^{-1}, \quad j = 1, \dots, 2^{p+q}, \quad (4.12)$$

where X_j denotes a real square root of J_R and Θ is an arbitrary nonsingular symplectic matrix which commutes with $\operatorname{diag}(J_R, J_R^T)$.

- (2) *If W has some real negative eigenvalues, then W has real Hamiltonian square roots which are not functions of W and they form 2^{p+q} parameterized families given by*

$$Y_j(\Theta) = \Psi \Theta \begin{bmatrix} \hat{X}_j & K_j \\ \hat{K}_j & -\hat{X}_j^T \end{bmatrix} \Theta^{-1} \Psi^{-1}, \quad j = 1, \dots, 2^{p+q}, \quad (4.13)$$

where \hat{X}_j is a square root for the Jordan blocks of J_R which are not associated with real negative eigenvalues, K_j and \hat{K}_j are symmetric block diagonal matrices corresponding to the square roots of the real negative eigenvalues, and Θ is an arbitrary nonsingular symplectic matrix which commutes with $\operatorname{diag}(J_R, J_R^T)$.

Proof. Equation (4.12) is a special case of Equation (4.11) in Theorem 4.5. If X_j is a real square root of J_R then $\operatorname{diag}(X_j, -X_j^T)$ is a Hamiltonian square root of $\operatorname{diag}(J_R, J_R^T)$ and Hamiltonian structure is preserved under symplectic similarity transformations. There are 2^{p+q} real square roots of J_R which may be or not functions of J_R .

For the second part, assume that J_R in (2.17),

$$J_R = \begin{bmatrix} C_{n_1}(a_1, b_1) & & & & & \\ & \ddots & & & & \\ & & C_{n_p}(a_p, b_p) & & & \\ & & & J_{n_{p+1}}(\lambda_{p+1}) & & \\ & & & & \ddots & \\ & & & & & J_{n_{p+q}}(\lambda_{p+q}) \end{bmatrix},$$

has only one real negative eigenvalue, say $\lambda_k < 0$, $k > p$ corresponding to the real Jordan block J_{n_k} .

Let $\pm iM_{n_k}$ with $M_{n_k} \in \mathbb{R}^{n \times n}$ be the two pure imaginary square roots of J_{n_k} , which are upper triangular Toeplitz matrices. See Corollary 4.1 and (2.3). Observe that $(\pm iM_{n_k})^2 = -M_{n_k}^2 = J_{n_k}$. We will first construct a square root of $\text{diag}(J_{n_k}, J_{n_k}^T)$ which is real and Hamiltonian. Let P_{n_k} be the reversal matrix of order n_k which satisfies $P_{n_k}^2 = I$ (the anti-diagonal entries are all 1's, the only nonzero entries). The matrices $P_{n_k}M_{n_k}$ and $M_{n_k}P_{n_k}$ are real symmetric and we have

$$\begin{bmatrix} & M_{n_k}P_{n_k} \\ -P_{n_k}M_{n_k} & \end{bmatrix}^2 = \begin{bmatrix} -M_{n_k}^2 & \\ & -P_{n_k}M_{n_k}^2P_{n_k} \end{bmatrix} = \begin{bmatrix} J_{n_k} & \\ & J_{n_k}^T \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} & M_{n_k}P_{n_k} \\ -P_{n_k}M_{n_k} & \end{bmatrix} \quad (\text{ and also } \begin{bmatrix} & -M_{n_k}P_{n_k} \\ P_{n_k}M_{n_k} & \end{bmatrix})$$

is a real Hamiltonian square root of $\text{diag}(J_{n_k}, J_{n_k}^T)$ which is not a function of $\text{diag}(J_{n_k}, J_{n_k}^T)$.

If $X_1 = \text{diag}(F_{n_1}, \dots, F_{n_p})$ is a real square root of $\text{diag}(C_{n_1}, \dots, C_{n_p})$, $X_2 = \text{diag}(L_{n_{p+1}}, \dots, L_{n_{k-1}})$ is a real square root of $\text{diag}(J_{n_{p+1}}, \dots, J_{n_{k-1}})$ and $X_3 = \text{diag}(L_{n_{k+1}}, \dots, L_{n_{p+q}})$ is a real square root of $\text{diag}(J_{n_{k+1}}, \dots, J_{n_{p+q}})$, then

$$\begin{bmatrix} X_1 & & & & \\ & X_2 & & & \\ & & O & & \\ & & & X_3 & \\ & & & & -X_1^T \\ & & & & & -X_2^T \\ & & & & & & O \\ & & & & & & & -X_3^T \\ & & & & & & & & -P_{n_k}M_{n_k} \end{bmatrix} =: \begin{bmatrix} \hat{X}_j & K_j \\ \hat{K}_j & -\hat{X}_j^T \end{bmatrix}$$

is a real Hamiltonian square root of $\text{diag}(J_R, J_R^T)$. Notice that there are 2^{p+q} different square roots with this form. Thus, for an arbitrary nonsingular symplectic matrix Θ which commutes with $\text{diag}(J_R, J_R^T)$,

$$Y_j(\Theta) = \Psi\Theta \begin{bmatrix} \hat{X}_j & K_j \\ \hat{K}_j & -\hat{X}_j^T \end{bmatrix} \Theta^{-1}\Psi^{-1}, \quad j = 1, \dots, 2^{p+q},$$

is a Hamiltonian square root of W .

If W has more than one real negative eigenvalue, the generalization is straightforward. \square

5. Algorithms for computing square roots of a skew-Hamiltonian matrix. In this section we will present a structure-exploiting Schur method to compute a real skew-Hamiltonian or a real Hamiltonian square root of a real skew-Hamiltonian matrix $W \in \mathbb{R}^{2n \times 2n}$ when W does not have real negative eigenvalues.

5.1. Skew-Hamiltonian square roots. First we obtain the *PVL* decomposition of $W \in \mathbb{R}^{n \times n}$ described in section 2.2,

$$U^T W U = \begin{bmatrix} W_1 & W_2 \\ O & W_1^T \end{bmatrix}, \quad W_2^T = -W_2,$$

where U is symplectic-orthogonal and W_1 is upper Hessenberg. The matrix U is constructed as a product of elementary symplectic-orthogonal matrices. These are the $2n \times 2n$ Givens rotations matrices of the type

$$\begin{bmatrix} I_{j-1} & & & \\ & \cos \theta & & \sin \theta \\ & & I_{n-1} & \\ & -\sin \theta & & \cos \theta \\ & & & & I_{n-j} \end{bmatrix}, \quad 1 \leq j \leq n,$$

for some angle $\theta \in [-\pi/2, \pi/2[$, and the direct sum of two identical $n \times n$ Householder matrices

$$H_j \oplus H_j(\mathbf{v}, \beta) = \begin{bmatrix} I_n - \beta \mathbf{v} \mathbf{v}^T & \\ & I_n - \beta \mathbf{v} \mathbf{v}^T \end{bmatrix},$$

where \mathbf{v} is a vector of length n with its first $j-1$ elements equal to zero. A simple combination of these transformations can be used to zero out entries in W to accomplish the PVL form. See Algorithm 1 and Algorithm 5 in [2, pp. 4,10]. The product of the transformations used in the reductions is accumulated to form the matrix U .

Then the standard QR algorithm is applied to W_1 producing an orthogonal matrix Q and a quasi-upper triangular matrix N_1 in real Schur form (2.8) so that

$$W_1 = Q N_1 Q^T,$$

and we attain the real skew-Hamiltonian Schur decomposition of W ,

$$\mathcal{T} = \mathcal{U}^T W \mathcal{U} = \begin{bmatrix} N_1 & N_2 \\ O & N_1^T \end{bmatrix}, \quad N_2 = -N_2^T,$$

where $\mathcal{U} = U \begin{bmatrix} Q & O \\ O & Q \end{bmatrix}$ and $N_2 = Q^T W_2 Q$.

This procedure takes only approximately a 20% of the computational cost the standard QR algorithm would require to compute the unstructured real Schur decomposition of W [2, p. 10].

Let

$$Z = \begin{bmatrix} X & Y \\ & X^T \end{bmatrix}, \quad Y = -Y^T.$$

be a skew-Hamiltonian square root of \mathcal{T} . We can solve the equation $Z^2 = \mathcal{T}$ exploiting the structure. From

$$\begin{bmatrix} X & Y \\ & X^T \end{bmatrix} \cdot \begin{bmatrix} X & Y \\ & X^T \end{bmatrix} = \begin{bmatrix} X^2 & XY + YX^T \\ 0 & (X^T)^2 \end{bmatrix} = \begin{bmatrix} N_1 & N_2 \\ 0 & N_1^T \end{bmatrix},$$

we have

$$X^2 = N_1 \tag{5.1}$$

and

$$XY + YX^T = N_2. \tag{5.2}$$

Equation (5.1) can be solved using Higham's real Schur method (see Algorithm 1, Section 3.3, page 12) and it is not difficult to show that X inherits N_1 's quasi-upper triangular structure. Equation (5.2) is a Lyapunov equation which can be solved efficiently since X is already in quasi-upper triangular real Schur form and Y is skew-symmetric. The techniques are the same as for the Sylvester equation. See [1, 14, Chapter 16].

If the partitions of $X = (X_{ij})$, $Y = (Y_{ij})$ and $N_2 = (N_{ij})$ are conformal with N_1 block structure,

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ & X_{22} & \dots & X_{2m} \\ & & \ddots & \vdots \\ & & & X_{mm} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_{11} & -Y_{21}^T & \dots & -Y_{m1}^T \\ Y_{21} & Y_{22} & \dots & -Y_{m2}^T \\ \vdots & \vdots & \ddots & \vdots \\ Y_{m1} & Y_{m2} & \dots & Y_{mm} \end{bmatrix},$$

$$N_2 = \begin{bmatrix} N_{11} & -N_{21}^T & \dots & N_{m1}^T \\ N_{21} & N_{22} & \dots & -N_{m2}^T \\ \vdots & \vdots & \ddots & \vdots \\ N_{m1} & N_{m2} & \dots & N_{mm} \end{bmatrix}, \quad Y_{ii} = -Y_{ii}^T, \quad N_{ii} = -N_{ii}^T, \quad i = i, \dots, m.$$

then, from (5.2), we have

$$\sum_{k=i}^m X_{ik}Y_{kj} + \sum_{k=j}^m Y_{ik}X_{jk}^T = N_{ij} \quad (5.3)$$

and

$$\begin{aligned} X_{ii}Y_{ij} + Y_{ij}X_{jj}^T &= N_{ij} - \sum_{k=i+1}^m X_{ik}Y_{kj} - \sum_{k=j+1}^m Y_{ik}X_{jk}^T \\ &= N_{ij} - \sum_{k=i+1}^m X_{ik}Y_{kj} - \sum_{k=j+1}^m Y_{ik}X_{jk}^T. \end{aligned}$$

These equations may be solved successively for $Y_{mm}, Y_{m,m-1}, \dots, Y_{m1}, Y_{m-1,m-1}, Y_{m-1,m-2}, \dots, Y_{m-1,1}, \dots, Y_{22}, Y_{21}$ and Y_{11} . We have to solve

$$\begin{aligned} X_{ii}Y_{ij} + Y_{ij}X_{jj}^T &= N_{ij} - \sum_{k=i+1}^m X_{ik}Y_{kj} - \sum_{k=j+1}^i Y_{ik}X_{jk}^T + \sum_{k=i+1}^m Y_{ki}^T X_{jk}^T, \\ i &= m, m-1, \dots, 1 \\ j &= i, i-1, \dots, 1. \end{aligned} \quad (5.4)$$

Since X_{ii} are of order 1 or 2, each system (5.4) is a linear system of order 1, 2 or 4 and is usually solved by Gaussian elimination with complete pivoting. The solution is unique because X_{ii} and $-X_{jj}^T$ have no eigenvalues in common. See Section 3.2.

Algorithm 2 [Skew-Hamiltonian real Schur method]

1. compute a real skew-Hamiltonian Schur decomposition of W ,

$$\mathcal{T} = \mathcal{U}^T W \mathcal{U} = \begin{bmatrix} N_1 & N_2 \\ 0 & N_1^T \end{bmatrix};$$

2. use Algorithm 1 to compute a square root X of N_1 , $X^2 = N_1$;
3. solve the Sylvester equation $XY + YX^T = N_2$ using (5.4) and form

$$Z = \begin{bmatrix} X & Y \\ & X^T \end{bmatrix};$$

4. obtain the skew-Hamiltonian square root of N , $\mathcal{X} = \mathcal{U}Z\mathcal{U}^T$.

The cost of the real skew-Hamiltonian Schur method for $W \in \mathbb{R}^{2n \times 2n}$ is measured in flops as follows. The real skew-Hamiltonian Schur factorization of W costs about

$3(2n)^3$ flops [11, 2]. The computation of X requires $n^3/6$ flops, the computation of the skew-symmetric solution Y requires about n^3 flops [11, p. 368] and the formation of $\mathcal{X} = \mathcal{U}Z\mathcal{U}^T$ requires $3(2n)^3/2$ flops. The total cost is approximately $5(2n)^3$ flops. Comparing with the overall cost of Algorithm 1, the unstructured real Schur method, which is about $17 \times (2n)^3$ flops, Algorithm 2 requires considerably less floating point operations.

5.2. Hamiltonian square roots. Analogously, let Z be a Hamiltonian square root of \mathcal{T} ,

$$Z = \begin{bmatrix} X & Y \\ & -X^T \end{bmatrix}, \quad Y = Y^T$$

(which is not a function of \mathcal{T}). To solve the equation $Z^2 = \mathcal{T}$, observe that, from

$$\begin{bmatrix} X & Y \\ & -X^T \end{bmatrix} \cdot \begin{bmatrix} X & Y \\ & -X^T \end{bmatrix} = \begin{bmatrix} X^2 & XY - YX^T \\ 0 & (X^T)^2 \end{bmatrix} = \begin{bmatrix} N_1 & N_2 \\ 0 & N_1^T \end{bmatrix}$$

it follows

$$X^2 = N_1 \tag{5.5}$$

and

$$XY - YX^T = N_2. \tag{5.6}$$

Equation (5.5) can be solved using Higham's real Schur method and Equation (5.6) is a singular Sylvester equation with infinitely many symmetric solutions. See [8, Proposition 7]. Again, the structure can be exploited and we have to solve

$$\begin{aligned} X_{ii}Y_{ij} - Y_{ij}X_{jj}^T &= N_{ij} - \sum_{k=i+1}^m X_{ik}Y_{kj} + \sum_{k=j+1}^i Y_{ik}X_{jk}^T - \sum_{k=i+1}^m Y_{ki}^T X_{jk}^T, \\ i &= m, m-1, \dots, 1 \\ j &= i, i-1, \dots, 1. \end{aligned} \tag{5.7}$$

The solution of the linear system (5.7) may not be unique but it always exists.

Algorithm 3 [Hamiltonian real Schur method]

1. compute a real skew-Hamiltonian Schur decomposition of W ,

$$\mathcal{T} = \mathcal{U}^T W \mathcal{U} = \begin{bmatrix} N_1 & N_2 \\ 0 & N_1^T \end{bmatrix};$$

2. use Algorithm 1 to compute a square root X of N_1 , $X^2 = N_1$;
3. obtain one solution for the Sylvester equation $XY - YX^T = N_2$ using (5.7) and form

$$Z = \begin{bmatrix} X & Y \\ & -X^T \end{bmatrix};$$

4. obtain the Hamiltonian square root of W , $\mathcal{X} = \mathcal{U}Z\mathcal{U}^T$.

6. Numerical examples. We implemented Algorithms 2 and 3 in MATLAB 7.5.0342 (R2007b) and used the Matrix Function Toolbox by Nick Higham available in Matlab Central website <http://www.mathworks.com/matlabcentral>. To find the square root X in step 2 we used the function `sqrtn_real` of this toolbox and to solve the linear systems (5.4) in step 3 of Algorithm 2 we used the function `sylvsol` (the solution is always unique). In step 3 of Algorithm 3 the linear systems (5.7) are solved using Matlab's function `pinv` which produces the solution with the smallest norm when the system has infinitely many solutions.

Let $\bar{\mathcal{X}}$ be an approximation to a square root of W and define the *residual*

$$E = \bar{\mathcal{X}}^2 - W.$$

Then, we have $\bar{\mathcal{X}}^2 = W + E$ and, as observed by Higham [13, p. 418], the stability of an algorithm for computing a square root $\bar{\mathcal{X}}$ of W corresponds to the residual E being small relative to W . Furthermore, for $\bar{\mathcal{X}}$ computed with `sqrtn_real`, Higham gives the following error bound

$$\frac{\|E\|_F}{\|W\|_F} \leq \left(1 + cn \frac{\|\bar{\mathcal{X}}\|_F^2}{\|W\|_F}\right) u$$

where $\|\cdot\|_F$ is the Frobenius norm, c is a constant of order 1, n is the dimension of W and u is the roundoff unit. Therefore, the real Schur method is stable provided that the number

$$\alpha(\mathcal{X}) = \frac{\|\bar{\mathcal{X}}\|_F^2}{\|W\|_F}$$

is small.

We expect our structure-preserving algorithms, Algorithm 2 (skew-Hamiltonian square root) and Algorithm 3 (Hamiltonian square root) to be as accurate as Algorithm 1 (real Schur method) which ignores the structure. The numerical examples

that follow illustrate that the three algorithms are all quite accurate when $\alpha(\mathcal{X})$ is small.

EXAMPLE 6.1. *The skew-Hamiltonian matrix*

$$W = \begin{bmatrix} \mathbf{e}\mathbf{e}^T & A \\ -A^T & \mathbf{e}\mathbf{e}^T \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 10^{-6} & 1 & 0 & 0 \\ -10^{-6} & 0 & 1 & 10^{-6} & 0 \\ -1 & -1 & 0 & 10^{-6} & 1 \\ 0 & -10^{-6} & -10^{-6} & 0 & 1 \\ 0 & 0 & -1 & -1 & 0 \end{bmatrix},$$

where \mathbf{e} is the vector of all ones, has one complex conjugate eigenvalue pair and 3 positive real eigenvalues (all with multiplicity 2).

The relative residuals of both the skew-Hamiltonian and Hamiltonian square roots computed with Algorithm 2 and Algorithm 3 are 4×10^{-15} , the same as for the square root delivered by Algorithm 1.

EXAMPLE 6.2. *The eigenvalues of the skew-Hamiltonian matrix*

$$W = \begin{bmatrix} A & B \\ B & A^T \end{bmatrix}, \quad A = \begin{bmatrix} 0 & -10^{-6} & 0 & 0 \\ 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 10^{-6} \\ 0 & 0 & -10^{-6} & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 1 & 2 & 3 \\ -1 & 0 & 2 & 3 \\ -2 & -2 & 0 & 3 \\ -3 & -3 & -3 & 0 \end{bmatrix},$$

are all very close to pure imaginary (four distinct eigenvalues).

The relative residuals of the square roots delivered by all the three methods are 4×10^{-16} .

If W has negative real eigenvalues there are no real square roots which are functions of W . However, all these algorithms can be applied and complex square roots will be obtained. In step 2 of Algorithms 2 and 3 a complex square root is computed and so we get a complex skew-Hamiltonian and a complex Hamiltonian square-root.

EXAMPLE 6.3.

For random matrices A , B and C (values drawn from a uniform distribution on the unit interval), the computed square roots of the skew-hamiltonian matrix of order $2n = 50$ (several cases)

$$W = \begin{bmatrix} A & B - B^T \\ C - C^T & A^T \end{bmatrix}$$

also have relative residuals of order at most 10^{-14} .

7. Conclusions. Based in the real skew-Hamiltonian Jordan form, we gave a clear characterization of the square roots of a real skew-Hamiltonian matrix W . This includes those square roots which are functions of W and those which are not. Although the Jordan canonical form is the main theoretical tool in our analysis, it is not suitable for numerical computation. We have designed a method for the computation of square roots of such structured matrices. An important component of our method is the real Schur decomposition tailored for skew-Hamiltonian matrices, which has been used by others in solving problems different from ours.

Our algorithm requires considerably less floating point operations (about 70% less) than the general real Schur method due to Higham. Furthermore, in numerical experiments, our algorithm has produced results which are as accurate as those obtained with `sqrtm_real`.

REFERENCES

- [1] R. H. Bartels and G. W. Stewart. Algorithm 432: Solution of the Matrix Equation $AX+XB=C$. *Communications of the ACM*, 15(9): 820–826, 1972.
- [2] P. Benner and D. Kressner. Fortran 77 Subroutines for Computing the Eigenvalues of Hamiltonian Matrices II. *ACM Transactions on Mathematical Software*, Vol. V, No. N: 1–22, 2005.
- [3] A. Bunse-Gerstner, R. Byers, and V. Mehrmann. A chart of numerical methods for structured eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 13: 419–453, 1993.
- [4] A. Björck and S. Hammarling. A Schur method for the square root of a matrix. *Linear Algebra and its Applications*, 52/53: 127–140, 1983.
- [5] S. Bora and V. Mehrmann. Linear perturbation theory for structured matrix pencils arising in control theory. *SIAM J. Matrix Analysis and Applications*, 28: 148–169, 2006.
- [6] G.W. Cross and P. Lancaster. Square roots of complex matrices. *Linear and Multi-linear Algebra*, 1:289–293, 1974.
- [7] P. I. Davis and N. J. Higham. A Parlett-Schur algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25:464–485, 2003.
- [8] H. Faßbender, S.D. Mackey, N. Mackey and H. Xu. Hamiltonian square roots of skew-Hamiltonian matrices. *Linear Algebra and its Applications*, 287: 125–159, 1999.
- [9] H. Faßbender and K. Ikramov. Several observations on symplectic, Hamiltonian, and skew Hamiltonian matrices. *Linear Algebra and its Applications*, 400:15–29,2005.
- [10] F. R. Gantmacher. *The Theory of Matrices*. Vol. 1, Chelsea, New York, 1977.
- [11] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore and London, 3rd edition, 1996.
- [12] N. J. Higham. Computing real square roots of a real matrix. *Linear Algebra and its Applications*, 88/89: 405–430, 1987.
- [13] N. J. Higham. Stable iterations for the matrix square root. *Numerical Algorithms*, 15:227–242, 1997.
- [14] N. J. Higham. *Accuracy and stability of numerical algorithms*. SIAM, second edition, 2002.
- [15] N. J. Higham. *Functions of matrices, theory and computation*. SIAM, 2008.
- [16] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur. Function preserving matrix groups

- and iterations for the matrix square root. *SIAM J. Matrix Analysis and Applications*, 26(3):849–877, 2005.
- [17] N. J. Higham. *Functions of matrices, theory and computation*. SIAM, 2008.
 - [18] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.
 - [19] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1991.
 - [20] P. Lancaster. *Theory of Matrices*. Academic, New York, 1969.
 - [21] L. Lin and Z.Y. Liu. On the square root of an h -matrix with positive diagonal elements. *Annals of Operations Research*, 103:339–350, 2001.
 - [22] Z. Liu, H. Chen, and H. Cao. The computation of the principal square roots of centrosymmetric h -matrices. *Applied Mathematics and Computation*, 175(1):319–329, 2006.
 - [23] Z. Liu, Y. Zhang and R. Ralha. Computing the square roots of matrices with central symmetry. *Applied Mathematics and Computation*, 186(1):715–726, 2007.
 - [24] C. Van Loan. A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix. *Linear Algebra and its Applications*, 61:233–251, 1984.
 - [25] D. S. Mackey, N. Mackey and F. Tisseur. Structured tools for structured matrices. *Electronic Journal of Linear Algebra*, 10:106–145, 2003.
 - [26] C. C. Paige and C. F. Van Loan. A Schur decomposition for Hamiltonian matrices. *Linear Algebra and its Applications*, 41:11–32, 1981.
 - [27] M. I. Smith. A Schur algorithm for computing matrix p th roots. *SIAM J. Matrix Analysis and Applications*, 24:971–989, 2003.